FEBS Letters 554 (2003) 410–416

FEBS 27818

Neuropsin (Opn5): a novel opsin identified in mammalian neural tissue

Emma E. Tarttelin[a,b,*], James Bellingham[a,c], Mark W. Hankins[a], Russell G. Foster[a], Robert J. Lucas[a,b]

[a] *Department of Integrative and Molecular Neuroscience, Imperial College London, Charing Cross Hospital, St. Dunstan's Road, London W6 8RP, UK*
[b] *School of Biological Sciences, University of Manchester, 1.124 Stopford Building, Oxford Road, Manchester M13 9PT, UK*
[c] *Gene Targeting Unit, Department of Neuromuscular Diseases, Division of Neuroscience and Psychological Medicine, Faculty of Medicine, Imperial College London, Charing Cross Hospital, St. Dunstan's Road, London W6 8RP, UK*

Received 18 September 2003; accepted 18 October 2003

First published online 27 October 2003

Edited by Ned Mantei

Abstract **We have cloned and characterised the expression of a new opsin gene, neuropsin (Opn5), in mice and humans. Neuropsin comprises seven exons on mouse chromosome 17. Its deduced protein sequence suggests a polypeptide of 377 amino acids in mice (354 in humans), with many structural features common to all opsins, including a lysine in the seventh transmembrane domain required to form a Schiff base link with retinaldehyde. Neuropsin shares 25–30% amino acid identity with all known opsins, making it the founding member of a new opsin family. It is expressed in the eye, brain, testis and spinal cord.**

Key words: Opsin; Photopigment; Photoisomerase; Neural expression

1. Introduction

The opsins are a superfamily of membrane bound heptahelical G-protein coupled receptors that can be distinguished on the basis of their amino acid sequence similarity and by the presence of a lysine (K) residue located in the seventh transmembrane domain [1]. During the course of vertebrate evolution gene duplications have led to the formation of distinct opsin families. Within these families, members exhibit $>75\%$ amino acid identity, even from distantly related species such as human and goldfish, whilst opsins from different families show 25–40% identity to one another [2].

Despite the diversity of opsin families, all those members for which a biochemical function has been described use the conserved lysine residue in the seventh transmembrane domain to bind a retinaldehyde chromophore via a Schiff base linkage. The photosensitivity of retinaldehyde is critical to their function. Members of the vertebrate ancient (VA)-opsin [3], pinopsin [4], as well as the rod and cone opsin families [1] employ the 11-*cis* isoform of retinaldehyde to function as sensory photopigments. Absorption of light by the 11-*cis* chromophore causes an isomerisation to the all-*trans* conformation, which induces conformational changes in the opsin protein, allowing activation of a G-protein phototransduction cascade [5]. The other opsin families with a clearly defined function (peropsin and RGR-opsin) have a different, but related role as photoisomerases. They bind all-*trans* retinaldehyde and use light energy to regenerate the 11-*cis* isoform [6–9].

The mammalian genome lacks several of the opsin families retaining only members of the rod opsin, some cone opsin, melanopsin, encephalopsin/panopsin, RGR-opsin and peropsin families. Two of these, melanopsin and encephalopsin, have unknown function. The inner retina of mammals contains recently discovered non-rod non-cone photoreceptors, which employ a retinaldehyde-based photopigment, whose identity remains unknown [10–13]. Currently, melanopsin and encephalopsin are the only opsins known to be present in the mammalian inner retina [14]. The best candidate for the protein moiety of the unidentified photopigment is melanopsin, which is expressed in these inner retinal photoreceptors [13]. Ablation of the melanopsin gene renders them insensitive to light [15]. Thus, melanopsin is a critical component of inner retinal photoreceptors. However, whether it functions as a photopigment and/or a photoisomerase remains unclear [16]. Inner retinal photoreceptors are distant from the retinal pigmented epithelium (RPE), the principle site of chromophore regeneration, and may require their own local isomerase. It may be that multiple opsin proteins are involved in the operation and maintenance of these novel photoreceptors.

An interesting feature of encephalopsin is its expression in cells or tissues classically considered to be non-photosensitive, such as lung, liver and kidney [17]. In teleosts a different opsin family (TMT-opsin) shows a similarly widespread expression pattern [18]. An expansion in assays for photoresponsiveness has revealed light sensitivity in many peripheral organs in fish [19], suggesting a role for TMT-opsins in photodetection. By contrast, the eyes and skin remain the only mammalian tissues known to be directly photosensitive. Whether the extraocular expression of mammalian encephalopsin reflects currently undiscovered photosensitivity in these tissues or a role for this opsin in non-photosensory pathways remains unknown.

In this paper we describe a bioinformatic and molecular approach for the identification of new members of the opsin superfamily in the mammalian genome. We report the isolation of a novel mammalian opsin-like gene, *Opn5*, the first member of a new opsin family. We have determined its gene

*Corresponding author. Fax: (44)-161-275 5363.
E-mail address: e.tarttelin@imperial.ac.uk (E.E. Tarttelin).

[1] Nucleotide sequence data reported are available in GenBank database under the accession numbers AY318865 and BK001605 (mouse) and AY377391 (human).

0014-5793/03/$22.00
doi:10.1016/S0014-5793(03)01212-2

structure, chromosomal localisation and tissue expression pattern in mice and humans. The expression of *Opn5* in the eye (retina and RPE) and brain suggests that it will provide new opportunities to examine the possible roles of mammalian opsins in inner retinal photoreception and/or non-light-dependent pathways.

2. Materials and methods

2.1. Database searches

A bioinformatic screen for novel opsins was undertaken using the TBLASTN programme (http://www.ncbi.nlm.nih.gov/BLAST) [20], using the protein coding sequences of known opsins (rhodopsin, XM_003284; pinopsin, AAA64223; peropsin, AAC51757; RGR-opsin, AAA56748; and melanopsin, NP_150598) as query sequences to look for significant matches, including a match to the critical lysine residue (equivalent to K296 of rhodopsin). Default parameters were used with the descriptions and alignments options set at 250. The BLAST suite of programmes at NCBI (http://www.ncbi.nlm.nih.gov/BLAST) was also used for subsequent screening and sequence data analysis.

2.2. RNA isolation and tissue expression

Total RNA was isolated from a panel of nine mouse tissues, human retina and brain, and the human Y79 and ARPE19 cell lines [21,22] using TRI reagent (Sigma) according to the manufacturer's instructions. Single-stranded cDNA was synthesised using the SuperScript First-Strand Synthesis System for reverse transcription-polymerase chain reaction (RT-PCR) (Invitrogen). Primers were designed to amplify a 218 bp fragment (NOPF: 5′-AGC CTT TGG AAG GCC AGA C-3′ and NOPR: 5′-CAG CAC AGC AGA AGA CTT C-3′) in order to analyse the tissue expression pattern of the novel opsin. This product spanned an intron to avoid contamination from genomic DNA. *Hprt* (hypoxanthine phosphoribosyltransferase) was used as a control to demonstrate mRNA reverse transcription fidelity in all samples. The *Hprt* primers encompassed two introns and were designed to amplify a fragment of 334 bp: 5′-GAT GAA CCA GGT TAT GAC C-3′ and 5′-TTG AGA GAT CAT CTC CAC C-3′. PCRs were performed in a total volume of 25 μl with 12.5 pmol of each primer, 0.5 nmol each of deoxyadenosine triphosphate (dATP), deoxycytidine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP) and deoxythymidine triphosphate (dTTP), and 0.25 units of BioTaq polymerase (Bioline) in the manufacturer's NH_4 buffer at an annealing temperature of 55°C and at 1.5 mM $MgCl_2$.

2.3. 3′ end sequence

The 3′ end of the murine *Opn5* cDNA was obtained using the following gene specific primers in conjunction with the 3′ RACE System (Life Technologies) – NOPF and NOP2F: 5′-GAA GTC TTC TGC TGT GCT G-3′. The 3′ end of the human *Opn5* cDNA was obtained by a combination of sequencing and RT-PCR. IMAGE [23] clone 4824978 was obtained from the MRC Geneservice (Babraham, UK). This testis-derived clone was sequenced and found to contain the 3′ untranslated region (UTR) sequence of human *Opn5*. RT-PCR using the primers HNO3F: 5′-GCT GCA CAC CGT AAC CAC-3′ and HNO3R: 5′-CAG TTT CTG CCT ACA TGG-3′ confirmed the presence of this sequence in the human retina and brain *Opn5* transcripts.

2.4. In silico analysis to obtain the 5′ end sequence

An in silico approach was undertaken to identify the 5′ end sequence of the *Opn5* cDNA as conventional 5′ RACE proved unsuccessful. The partial murine *Opn5* sequence was used to screen the genomes/fugu_gs database using the BLASTX programme. A significant match to a predicted fugu protein was identified, and this protein was subsequently used to re-screen the databases with the TBLASTN programme. This enabled the identification of the orthologous regions of coding sequence within the mouse and human genomes. From this new sequence, primers were designed for use in RT-PCRs, to confirm the presence of the 5′ exons as part of the full-length mouse and human *Opn5* transcripts. The primers used were: RN1F: 5′-ACA GCA TGG CCT TGA ACC-3′, RN2F: 5′-ATT CAC GGA CTT GTT ATT CC-3′, HUMF1: 5′-ACA GAA TGG CGT TAA ATC AC-3′, NOPR and MNO6R: 5′-CTT ATT TCT GAA GCC CTG-3′.

2.5. Cloning and sequencing

All PCR products were ligated into pGEM-T-Easy vector (Promega) and transformed in DH5α subcloning efficiency competent cells (Invitrogen). Nucleotide sequence determination was carried out on an ABI Prism 377 DNA Sequencer (Perkin Elmer) using the ABI Prism BigDye terminator cycle sequencing kit (Perkin Elmer). To eliminate the possibility of PCR generated sequence errors the cloned products from three independent PCR amplifications were sequenced on both DNA strands.

2.6. Phylogenetic analysis

Nucleotide coding sequences were entered in to DAMBE version 4.1.15 [24] (http://aix1.uottawa.ca/~xxia/software/software.htm), and converted to amino acid sequences which were then aligned using the Clustal W [25] option. Maximum likelihood analysis was performed by quartet analysis using TREE-PUZZLE 5.0 [26] (http://www.tree-puzzle.de/). Reliability values are based on quartet puzzling with 10 000 puzzling steps, VT model and assuming uniform rate heterogeneity.

3. Results and discussion

3.1. Identification and isolation of the Opn5 gene in mouse and human

A bioinformatics approach was used to identify novel opsin-like sequences in the mammalian genomes that are available on the public databases. The GenBank database was screened with representative members of the various opsin classes using the TBLASTN algorithm. Only those unique results showing a match to the lysine residue that forms the retinal attachment site of the opsin query sequence were selected for further analysis. The only novel match obtained was for an expressed sequence tag (EST) (AI810121) corresponding to the 3′ end of IMAGE clone 2360502 which is derived from a pooled human library comprising foetal lung, B-cell and testis tissue. A second human EST (BG721121) corresponding to the 5′ end of IMAGE clone 4824978 that is derived from a human testis library was also identified in all of the database searches. The EST AI810121 sequence was subsequently used to search the databases through the TBLASTN programme. A significant match was identified to EST BB633508, corresponding to the 5′ end of a RIKEN mouse clone [27] from an adult spinal cord library. The available nucleotide sequence from these ESTs was used to design primers for further analysis.

Both RT-PCR and RACE PCR were carried out in order to isolate a partial cDNA sequence from mouse eye and brain. An in silico approach was then undertaken to complete the full-length coding sequences. The partial murine sequence identified a significant match of 71% identity to the predicted *Fugu* protein FuguGenscan_19252, comprising 359 amino acids. This protein sequence was subsequently used to re-screen the databases to identify the orthologous upstream sequences within the mouse genome. The putative 5′ exons were then confirmed by RT-PCR on mouse eye and brain cDNAs. This combination of approaches enabled the identification of a novel mouse cDNA sequence of 1825 bp (GenBank accession number: AY318865). This sequence contains an open reading frame of 1134 bp that encodes for a protein of 377 amino acids. BLAST analysis of the protein sequences confirmed that it is a member of the opsin superfamily, and no matches were observed to any other class of G-protein coupled receptors. The predicted amino acid sequence shows around 25–30% identity to other vertebrate and invertebrate members of the opsin superfamily, as shown in Table 1, defin-

Table 1
Amino acid identity (%) of the transmembrane domains (equivalent to amino acids 35–306 of the rod opsin model of Palczewski et al. [48]) shared by representatives of various opsin families

	RHO	P	VA	OPN3	RGR	RRH	OPN4	Opn5	*Sepia*
Human rod (RHO)	−/−								
Chicken pinopsin (P)	48	−/−							
Carp VA-opsin (VA)	37	45	−/−						
Human encephalopsin (OPN3)	31	30	29	−/−					
Human RGR-opsin (RGR)	23	23	21	25	−/−				
Human peropsin (RRH)	27	26	27	30	26	−/−			
Human melanopsin (OPN4)	27	28	27	28	26	30	−/−		
Mouse neuropsin (Opn5)	25	26	25	27	25	29	28	−/−	
Sepia rhodopsin (invertebrate)	27	26	23	25	22	27	41	30	−/−

ing it as a novel opsin family. For this reason we have designated the gene *Opn5* (opsin5). The putative mouse Opn5 sequence was analysed using the online TMHMM Server version 2.0 [28,29] (http://www.cbs.dtu.dk/services/TMHMM/) which predicted seven transmembrane domains as would be expected of an opsin protein (data not shown).

Subsequently, RT-PCR and sequence data from IMAGE clone 4824978 were used to identify the human *Opn5* gene. The human sequence comprises 1778 bp (GenBank accession number: AY377391) and contains an open reading frame of 1065 bp that encodes for a protein of 354 amino acids. Human and mouse Opn5 show high levels of identity to one another (Fig. 1A); 89% at the nucleotide level and 95% at the protein level, which suggests strict functional constraints. The most significant difference between the mouse and human sequences is at the C-terminal tail of the protein, where there

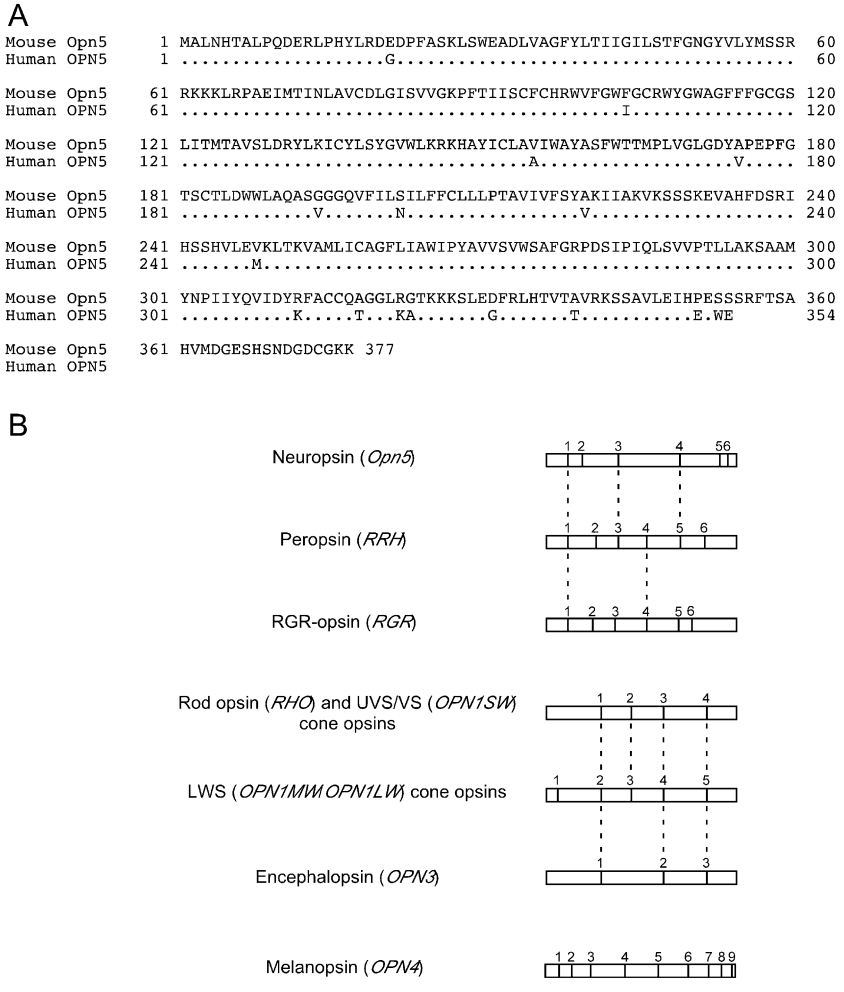

Fig. 1. A: Alignment of mouse and human Opn5 protein sequences. Dots indicate amino acid identity, which is 95% between the two species. B: Schematic diagram of the relative positions of introns of the opsin gene superfamily in mammalian genomes (numbered vertical bars). Conserved intron positions are indicated by dashed vertical lines. In the *Opn5* gene, the position of intron 1 is conserved with the corresponding intron in the peropsin gene, *RRH*, and the RGR-opsin gene, *RGR*. Intron 3 of *Opn5* is also conserved with intron 3 of *RRH*, whilst intron 4 of *Opn5* is at a position conserved with intron 5 of *RRH*. This conservation of intron positions may be indicative of a common ancestry [41].

are an additional 23 amino acids in mouse Opn5. The functional significance of the longer protein remains to be determined, but species differences in the length of the C-terminal tail of opsin proteins have been previously documented [30].

Alignment of the Opn5 proteins with representatives of other opsin classes reveals that several key features are conserved (see Fig. 2). These include the lysine residue at position 296 required to form the Schiff base with the chromophore, equivalent to K296 in rhodopsin. The positively charged Schiff base is balanced by a counterion in the third transmembrane domain, which is an acidic residue (glutamate) in most vertebrate visual opsins [2]. The putative counterion in the Opn5 protein is a tyrosine residue at position 109. Tyrosine has also been identified as the possible counterion in peropsin [31], melanopsin [32] and TMT-opsin [18], and is found in invertebrate visual pigments at this position [33]. Another characteristic feature of opsin proteins is a C-terminal tail enriched for serine and threonine residues as these are sites for phosphorylation by various kinases [34]. The C-terminal tail of both mouse and human Opn5 shows enrichment of these amino acids, and a scan of the sequence using the PROSITE programme [35] (http://us.expasy.org/cgi-bin/scanprosite) identified five such phosphorylation motifs. Additionally Opn5 contains an N-linked glycosylation site (residues 4–7), a site at the C-terminus for palmitoylation (C315, C316), two cysteine residues for disulphide bond formation (C106 and C183), and a conserved amino acid tripeptide aspartate, arginine and tyrosine (DRY 130–132) at the end of transmem-

```
Human Rod Opsin       1               MNGTEGPNFYVPFSNA--------------TGVVRSPFEYPQY-YLAEPWQFSMLAA  42
Chicken Pinopsin      1             MSSNSSQAP----PNG-------------------TPGPFDGPQWPYQAPQSTYVGVAV  36
Human Encephalopsin   1             MYSGNRSGGHGYWDGGGAA--------------GAEGPAPAGTLSPAPLFSPGTYERLAL  46
Human Peropsin        1             MLRNNLGNS----------------------------SDSKNEDG-SVFSQTEHNIVAT  30
Human RGR-opsin       1               MAETSALP-------------------------------------TGFGELEVLAVGM  21
Human Melanopsin      1 MNPPSGPRVPPSPTQEPSCMATPAPPSWWDSSQSSISSLGRLPSISPTAPGTWAAAWVPLPTVDVPDHAHYTLGT  75
Mouse Neuropsin       1             MALNHTALP----QDER-----------------LPHYLRDEDPFASKLSWEADLVAGF  38

                               TM 1                                  TM 2
Human Rod Opsin      43 YMFLLIVLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFMVLGGFTSTLYTSLHGYFVFGPTGCNLEGFFA 117
Chicken Pinopsin     37 LMGTVVACASVVNGLVIVVSICYKKLRSPLNYILVNLAVADLLVTLCGSSVSLSNNINGFFVFGRRMCELEGFMV 111
Human Encephalopsin  47 LLGSIGLLGVGNNLLVLVLYYKFQRLRTPTHLLLVNISLSDLLVSLFGVTFTFVSCLRNGWVWDTVGCVWDGFSG 121
Human Peropsin       31 YLIMAGMISIISNIIVLGIFIKYKELRTPTNAIIINLAVTDIGVSSIGYPMSAASDLYGSWKFGYAGCQVYAGLN 105
Human RGR-opsin      22 VLLVEALSGLSLNTLTIFSFCKTPELRTPCHLLVLSLALADSGISLN-ALVAATSSLLRRWPYGSDGCQAHGFQG  95
Human Melanopsin     76 VILLVGLTGMLGNLTVIYTFCRSRSLRTPANMFIINLAVSDFLMSFTQAPVFFTSSLYKQWLFGETGCEFYAFCG 150
Mouse Neuropsin      39 YLTIIGILSTFGNGYVLYMSSRRKKKLRPAEIMTINLAVCDLGISVVGKPFTIISCFCHRWVFGWFGCRWYGWAG 113
                                    *               *           *                          *
                             TM 3                                    TM 4
Human Rod Opsin     118 TLGGEIALWSLVVLAIERYVVVC-KPMSNFRFGE-NHAIMGVAFTWVMALACAAPPLAGWSRYIPEGLQCSCGID 190
Chicken Pinopsin    112 SLTGIVGLWSLAILALERYVVVC-RPLGDFQFQR-RHAVSGCAFTWGWALLWSTPPLLGWSSYVPEGLRTSCGPN 184
Human Encephalopsin 122 SLFGIVSIATLTVLAYERYIRVVHARVINF-----SWAWRAITYIWLYSLAWAGAPLLGWNRYILDVHGLGCTVD 191
Human Peropsin      106 IFFGMASIGLLTVVAVDRYLTIC-LPDVGRRMTT-NTYIGLILGAWINGLFWALMPIIGWASYAPDPTGATCTIN 178
Human RGR-opsin      96 FVTALASICSSAAIAWGRYHHYC----TRSQLAW-NSAVSLVLFVWLSSAFWAALPLLGWGHYDYEPLGTCCTLD 165
Human Melanopsin    151 ALFGISSMITLTAIALDRYLVIT-RPLATFGVASKRRAAFVLLGVWLYALAWSLPPFFGWSAYVPEGLLTSCSWD 224
Mouse Neuropsin     114 FFFGCGSLITMTAVSLDRYLKIC-YLSYGVWLKR-KHAYICLAVIWAYASFWTTMPLVGLGDYAPEPFGTSCTLD 186
                                         **                          *         *  *   *       *
                                                 TM 5
Human Rod Opsin     191 YYTLKPEVNNESFVIYMFVVHFTIPMIIIFFCYGQLVFTVKEA------------AAQQQESATTQKAEKEVTRM 253
Chicken Pinopsin    185 WYTGG--SNNNSYILSLFVTCFVLPLSLILFSYTNLLLTLRAA------------AAQQKEADTTQRAEREVTRM 245
Human Encephalopsin 192 WKSKD--ANDSSFVLFLFLGCLVVPLGVIAHCYGHILYSIRMLRCVE--------DLQTIQVIKILKYEKKLAKM 256
Human Peropsin      179 WRKND--RSFVSYTMTVIAINFIVPLTVMFYCYYHVTLSIKHHTT----------SDCTESLNRDWSDQIDVTKM 241
Human RGR-opsin     166 YSKGD--RNFTSFLFTMSFFNFAMPLFITITSYSLMEQKLGK--------------------------SGHLQVN 212
Human Melanopsin    225 YMSFT--PAVRAYTMLLCCFVFFLPLLIIIYCYIFIFRAIRETGRALQTFGACKGNGESLWQRQRLQSECKMAKI 297
Mouse Neuropsin     187 WWLAQASGGGQVFILSILFFCLLLPTAVIVFSYAKIIAKVKSSSKEV--------AHFDSRIHSSHVLEVKLTKV 253
                                               *       *
                             TM 6                                  TM 7
Human Rod Opsin     254 VIIMVIAFLICWVPYASVAFYIFTHQGSNFGPIFMTIPAFFAKSAAIYNPVIYIMMNKQFRNCMLTTICCG---- 324
Chicken Pinopsin    246 VIVMVMAFLLCWLPYSTFALVVATHKGIIIQPVLASLPSYFSKTATVYNPIIYVFMNKQFQSCLLEMLCCG---- 316
Human Encephalopsin 257 CFLMIFTFLVCWMPYIVICFLVVNGHGHLVTPTISIVSYLFAKSNTVYNPVIYVFMIRKFRRSLLQLLCLR---- 327
Human Peropsin      242 SVIMICMFLVAWSPYSIVCLWASFGDPKKIPPPMAIIAPLFAKSSTFYNPCIYVVANKKFRRAMLAMFKCQ---- 312
Human RGR-opsin     213 TTLPARTLLLGWGPYAILYLYAVIADVTSISPKLQMVPALIAKMVPTINAINYALGNEMVCRGIWQCLSPQ---- 283
Human Melanopsin    298 MLLVILLFVLSWAPYSAVALVAFAGYAHVLTPYMSSVPAVIAKASAIHNPIIYAITHPKYRVAIAQHLPCLGVLL 372
Mouse Neuropsin     254 AMLICAGFLIAWIPYAVVSVWSAFGRPDSIPIQLSVVPTLLAKSAAMYNPIIYQVIDYRFACCQAGGLRGT---- 324
                                  * **                             *     *   *

Human Rod Opsin     325 ------KNP-------LGDDEASATVSKTET------------SQVAPA                           348
Chicken Pinopsin    317 ------YQPQRTGKASPGTPGPHADVTAAGLR-----------NKVMPAHPV                        351
Human Encephalopsin 328 ------LLRCQRPAKDLPAAGSEMQIRPIVMSQK---------DGDRPKKKVTFNSSSIIFIITSDESLSVDDSD 387
Human Peropsin      313 ------THQ----------TMPVTSILPMDVSQ----------NPLASGRI                         337
Human RGR-opsin     284 --------------------KREKDRTK                                                291
Human Melanopsin    373 GVSRRHSRPYPSYRSTHRSTLTSHTSNLSWISIRRRQESLGSESEVGWTHMEAAAVWGAAQQANGRSLYGQGLED 447
Mouse Neuropsin     325 ------KKKS-------LEDFRLHTVTAVRKSS-----------AVLEIHPESSSRFTSAHVMDGESHSNDGDCG 375


Human Rod Opsin
Chicken Pinopsin
Human Encephalopsin 388 KTIGVQSLMLIQVRPL 403
Human Peropsin
Human RGR-opsin
Human Melanopsin    448 LEAKAPPRPQGHEAETPGKTKGLIPSQDPRM 478
Mouse Neuropsin     376 KK 377
```

Fig. 2. Alignment of the predicted amino acid sequence of neuropsin with representatives of other opsin groups indicates that several features of neuropsin are conserved such as the probable lysine retinal attachment site (K296) in the seventh transmembrane domain, and the potential tyrosine counterion (equivalent of E113) in the third transmembrane domain (numeration and helices based upon the rhodopsin model of Palczewski et al. [48]). Key features of neuropsin mentioned in the text are shown in bold or underlined.

brane domain 3, which is important for G-protein binding [36,37]. These features are found in many opsin classes, but are not diagnostic for the opsin family as they are also found in other G-protein coupled receptors [38].

The length of the second extracellular loop has been found to vary between opsins [4]. In rod and cone opsins this loop is two amino acids longer than in all other vertebrate and invertebrate opsins. Interestingly, like the rod and cone opsins, Opn5 has the extra two amino acids in the second extracellular loop between the fourth and fifth transmembrane domains (Fig. 2: alanine, serine AS 192–193). The second extracellular loop has been suggested to play a role in both the correct folding of the membrane embedded helices and in stabilisation of the physiologically active meta II intermediate form of the opsin molecule [39]. Site-directed mutagenesis studies have been carried out on this region of pinopsin, which has been shown to form a functional photopigment, but lacks the extra amino acid doublet [40]. Swapping a six amino acid stretch from extracellular loop 2 of chicken pinopsin for an equivalent eight amino acid stretch from this region of chicken green cone opsin resulted in a more rapid decay of the meta II intermediate of the mutant pinopsin. As the individual amino acid residues from this region are not conserved between different rod and cone pigments it is thought that the length of the loop sequence may be the factor important for determining the stability [40]. If Opn5 does form a photopigment then the presence of the extra doublet in extracellular loop 2 suggests that the turnover of the meta II Opn5 intermediate may be rapid.

3.2. *Gene structure and chromosomal localisation of* Opn5

A comparison of the mouse and human *Opn5* cDNA sequences with the genomic sequence available on the public database reveals that the gene comprises seven exons, with six introns present in the coding sequence (GenBank: BK001605). All introns are flanked by consensus acceptor and donor splice sites. Employing the model of Bellingham et al. [41], an alignment of the *Opn5* coding sequences with those of representatives from other opsin families indicates that the insertion site of intron 1 is conserved with the position of intron 1 in *RGR* (RGR-opsin) and *RRH* (peropsin). Introns 3 of *Opn5* and *RRH* are also in homologous positions, as are *Opn5* intron 4 and *RRH* intron 5. The positions of introns 2, 5 and 6 in the *Opn5* gene are unique amongst the known vertebrate opsin genes, confirming that Opn5 represents a novel opsin family (Fig. 1B).

In both mouse and human eye and brain a splice variant was detected that contained an extra exon between exons 3 and 4. However, this exon contains two in-frame stop codons, which would result in production of a truncated protein. Whether this alternative transcript has a functional role remains undetermined.

In order to localise the *Opn5* gene, the cDNA sequence was used as a search tool in a BLAST analysis of the murine genome. *Opn5* was found to map to mouse chromosome 17, to a region at approximately 40.8 Mb, corresponding to cytogenetic interval 17B3. This places *Opn5* between the known genes *Rhag* (rhesus blood group-associated A glycoprotein) at 39 Mb and *Cd2ap* (CD2-associated protein) at 41 Mb. Human *Opn5* maps to the orthologous region on the short arm of chromosome 6, to a region at approximately 47 Mb, corresponding to cytogenetic interval 6p21–p12.

Recently, a modifier locus was identified on mouse chromosome 17 that modulates acute light suppression of activity in melanopsin knockout animals depending on their genetic background [42]. Genotyping of the strains involved showed that the modifier gene is located in a 6 Mb region at approximately 19–25 Mb of mouse chromosome 17. This localisation excludes the *Opn5* gene from being the modifier for masking activity.

3.3. *Tissue expression of* Opn5 *(neuropsin)*

The expression pattern of the *Opn5* gene was analysed in a panel of nine mouse tissues. RT-PCR showed expression is detectable in the eye, brain and testis (Fig. 3A and B). Similarly, in humans, expression of *Opn5* was detected in the retina, brain and cell lines derived from neural retina (Y79) and the RPE (ARPE19) (Fig. 3C). Human *Opn5* is also expressed in the testis as ESTs AI810121 and BG721121 are both derived from testis cDNA libraries. The relevance of the testis expression is unclear. The finding that genes that are normally tissue specific can also be ectopically expressed in the testis has been previously documented. Retinal S-antigen (arrestin) for example, has been shown to be expressed in the testis [43], and we have observed expression of rod opsin, the rod cell photopigment in murine testis (data not shown). Whether Opn5 protein has a function in the testis or whether there is merely ectopic *Opn5* mRNA expression in the testis remains to be determined. Apart from the testis expression, *Opn5* expression appears to be neural specific, with known sites of expression in the eye, brain and spinal cord (EST BB633508). For this reason we have termed *Opn5* neuropsin (from *neu*ral *opsin*).

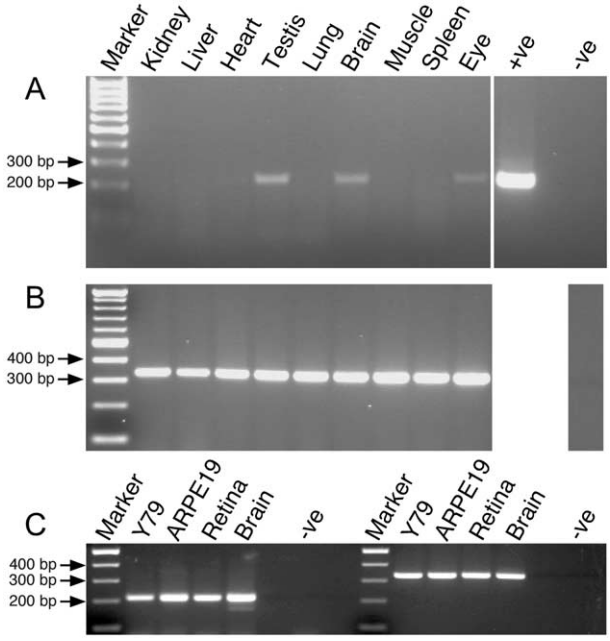

Fig. 3. RT-PCR analysis of the expression of neuropsin in a panel of mouse and human tissues. A: Neuropsin expression is observed in the mouse brain, eye and testis. The positive control (+ve) was a cloned and sequenced *Opn5* PCR product. Negative controls (−ve) had no cDNA. B: *Hprt* was used as the housekeeping positive control gene to confirm the fidelity of the reverse transcription reaction. C: Neuropsin (left side) and *HPRT* (right side) expression in human eye and brain. Neuropsin is expressed in the retina, brain and Y79 and ARPE19 cell lines.

Within the eye, human *Opn5* is expressed in both the neural retina (retinal and Y79 cDNA), and the RPE (ARPE19 cDNA). These sites of expression support the possibility that neuropsin may perform some function in light detection. In contrast, expression in the brain would appear to exclude such a role, as extraocular photoreception is not found in mammals. One possibility to exclude is that the RT-PCR signal from brain reflects *Opn5* expression only in the pineal gland, similar to the rod and cone opsins [44], as a portion of the neuropsin 3′ untranslated sequence identifies a mouse RIKEN EST (GenBank: AV381963), which is derived from an adult diencephalon library. In non-mammalian species selected areas within the diencephalon, in particular the anterior medial preoptic area and the paraventricular nuclei have been reported to be involved in light detection [45]. Although these brain regions have not been shown to play a role in light detection in mammals, there is evidence that light can permeate through the skull to areas of the mammalian brain including the diencephalon [46]. Encephalopsin (*Opn3*), another member of the opsin family, is expressed at high levels in several regions of the mouse brain, including the diencephalon [47]. Further characterisation of the sites of expression of neuropsin within the eye and brain will provide an insight into the potential role of this novel opsin.

3.4. Phylogenetic position of neuropsin

A maximum likelihood tree of representatives from various vertebrate opsin families was constructed from amino acid data (Fig. 4). It can be seen that neuropsin (Opn5) would appear to be most closely related to peropsin (RRH) in the RGR-opsin/peropsin clade. Although branch support levels for and within this clade are relatively low, and the levels of amino acid identity between Opn5 and RGR-opsin/peropsin are no higher than the levels of identity observed with the other opsin families, the phylogenetic relationship between these opsins is strengthened in part by their shared intron structure (see Section 3.2).

3.5. Conclusions

We have utilised both bioinformatic and molecular approaches to identify a new member of the opsin superfamily, neuropsin (*Opn5*) in the mammalian genome. Since neuropsin shows 25–30% identity to the known opsins, we have designated it the first member of a new opsin family. It retains all of the amino acid residues critical for opsin function including a putative lysine retinaldehyde chromophore binding site in the seventh transmembrane domain. Neuropsin expression in mouse is confined to the eye, brain, spinal cord and testes. Human neuropsin expression has been detected in brain and testis and has been sub-localised within the eye to the neural retina and RPE. Further sub-localisation of neuropsin expression within neural tissues, in conjunction with functional studies of the protein will provide more insight into its role. Our strategy of using both comparative bioinformatics and conventional molecular techniques was critical in identifying the full-length coding sequence for *Opn5*, and it is likely that similar approaches will be of use for the identification of rare transcripts in general. To date, our search strategy has failed to identify any further opsin-like genes in the mammalian genome. This suggests that if further opsin genes exist in mammals, they are currently not represented by ESTs and are in regions of the genome poorly represented on the databases.

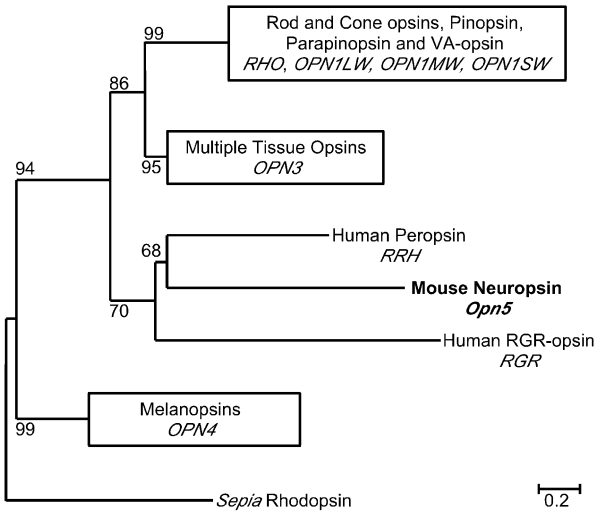

Fig. 4. Best maximum likelihood tree (log $L = -14408.49$) based on amino acid sequences of representatives of the extant vertebrate opsin classes showing the separation of mouse neuropsin (*Opn5*) with RGR-opsin and peropsin. Levels of branch support (%) are indicated, as are the recognised mammalian gene symbols (italicised). Major opsin groupings are collapsed to their common branch. The tree was rooted using an invertebrate opsin, *Sepia* (cuttlefish) rhodopsin. Scale bar: substitutions/site. Amino acid sequences used in tree construction are: human rod opsin, U49742; chicken green cone (MWS) opsin, M92038; chicken blue cone (SWS) opsin, M92037; human blue cone (UVS/VS) opsin, U53874; human red cone (LWS) opsin, M13305; carp VA-opsin, AF233520; catfish parapinopsin, AF028014; chicken pinopsin, U15762; human encephalopsin, AF140242; *Fugu* TMT-opsin, AF402774; mouse neuropsin, AY318865; human peropsin, AF012270; human RGR-opsin, U14910; human melanopsin, AF147788; zebrafish melanopsin, AY078161; *Xenopus* melanopsin, AF014797; *Sepia* rhodopsin, AF000947.

Acknowledgements: This work was supported by the BBSRC and The Hammersmith Hospitals Trust Special Trustees.

References

[1] Ebrey, T. and Koutalos, Y. (2001) Prog. Retin. Eye Res. 20, 49–94.
[2] Bellingham, J. and Foster, R.G. (2002) Cell Tissue Res. 309, 57–71.
[3] Soni, B.G., Philp, A.R., Foster, R.G. and Knox, B.E. (1998) Nature 394, 27–28.
[4] Okano, T., Yoshizawa, T. and Fukada, Y. (1994) Nature 372, 94–97.
[5] Burns, M.E. and Baylor, D.A. (2001) Annu. Rev. Neurosci. 24, 779–805.
[6] Koyanagi, M., Terakita, A., Kubokawa, K. and Shichida, Y. (2002) FEBS Lett. 531, 525–528.
[7] Chen, P. et al. (2001) Nat. Genet. 28, 256–260.
[8] Hao, W. and Fong, H.K.W. (1996) Biochemistry 35, 6251–6256.
[9] Hao, W. and Fong, H.K.W. (1999) J. Biol. Chem. 274, 6085–6090.
[10] Lucas, R.J., Douglas, R.H. and Foster, R.G. (2001) Nat. Neurosci. 4, 621–626.
[11] Hattar, S. et al. (2003) Nature 424, 75–81.
[12] Berson, D.M., Dunn, F.A. and Takao, M. (2002) Science 295, 1070–1073.
[13] Hattar, S., Liao, H.-W., Takao, M., Berson, D.M. and Yau, K.-W. (2002) Science 295, 1065–1070.
[14] Tarttelin, E.E. et al. (2003) Exp. Eye Res. 76, 393–396.
[15] Lucas, R.J., Hattar, S., Takao, M., Berson, D.M., Foster, R.G. and Yau, K.-W. (2003) Science 299, 245–247.
[16] Foster, R.G. and Bellingham, J. (2002) Curr. Biol. 12, R543–R544.

[17] Halford, S., Freedman, M.S., Bellingham, J., Inglis, S.L., Poopalasundaram, S., Soni, B.G., Foster, R.G. and Hunt, D.M. (2001) Genomics 72, 203–208.
[18] Moutsaki, P., Whitmore, D., Bellingham, J., Sakamoto, K., David-Gray, Z.K. and Foster, R.G. (2003) Mol. Brain Res. 112, 135–145.
[19] Whitmore, D., Foulkes, N.S. and Sassone-Corsi, P. (2000) Nature 404, 87–91.
[20] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[21] McFall, R.C., Sery, T.W. and Makadon, M. (1977) Cancer Res. 37, 1003–1010.
[22] Dunn, K.C., Aotaki-Keen, A.E., Putkey, F.R. and Hjelmeland, L.M. (1996) Exp. Eye Res. 62, 155–169.
[23] Lennon, G., Auffray, C., Polymeropoulos, M. and Soares, M.B. (1996) Genomics 33, 151–152.
[24] Xia, X. and Xie, Z. (2001) J. Hered. 92, 371–373.
[25] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[26] Schmidt, H.A., Strimmer, K., Vingron, M. and von Haeseler, A. (2002) Bioinformatics 18, 502–504.
[27] Shibata, K. et al. (2000) Genome Res. 10, 1757–1771.
[28] Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L.L. (2001) J. Mol. Biol. 305, 567–580.
[29] Sonnhammer, E.L.L., von Heijne, G. and Krogh, A. (1998) in: Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology (Glasgow, J., Littlejohn, T., Major, F., Lathrop, R., Sankoff, D. and Sensen, C., Eds.), pp. 175–182, AAAI Press, Menlo Park, CA.
[30] Bellingham, J., Whitmore, D., Philp, A.R., Wells, D.J. and Foster, R.G. (2002) Mol. Brain Res. 107, 128–136.
[31] Sun, H., Gilbert, D.J., Copeland, N.G., Jenkins, N.A. and Nathans, J. (1997) Proc. Natl. Acad. Sci. USA 94, 9893–9898.
[32] Provencio, I., Rodriguez, I.R., Jiang, G., Hayes, W.P., Moreira, E.F. and Rollag, M.D. (2000) J. Neurosci. 20, 600–605.
[33] Bellingham, J., Morris, A.G. and Hunt, D.M. (1998) J. Exp. Biol. 201, 2299–2306.
[34] Hurley, J.B., Spencer, M. and Niemi, G.A. (1998) Vis. Res. 38, 1341–1352.
[35] Sigrist, C.J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A. and Bucher, P. (2002) Brief Bioinform. 3, 265–274.
[36] Franke, R.R., König, B., Sakmar, T.P., Khorana, H.G. and Hofmann, K.P. (1990) Science 250, 123–125.
[37] Franke, R.R., Sakmar, T.P., Graham, R.M. and Khorana, H.G. (1992) J. Biol. Chem. 267, 14767–14774.
[38] Bockaert, J. and Pin, J.P. (1999) EMBO J. 18, 1723–1729.
[39] Doi, T., Molday, R.S. and Khorana, H.G. (1990) Proc. Natl. Acad. Sci. USA 87, 4991–4995.
[40] Nakamura, A., Kojima, D., Okano, T., Imai, H., Terakita, A., Shichida, Y. and Fukada, Y. (2001) J. Biochem. (Tokyo) 129, 329–334.
[41] Bellingham, J., Wells, D.J. and Foster, R.G. (2003) BMC Genomics 4, 3.
[42] Panda, S. et al. (2003) Science 301, 525–527.
[43] Derbinski, J., Schulte, A., Kyewski, B. and Klein, L. (2001) Nat. Immunol. 2, 1032–1039.
[44] Blackshaw, S. and Snyder, S.H. (1997) J. Neurosci. 17, 8074–8082.
[45] Foster, R.G., Grace, M.S., Provencio, I., Degrip, W.J. and García-Fernández, J.M. (1994) Neurosci. Biobehav. Rev. 18, 541–546.
[46] Muller, P.J. and Wilson, B.C. (1986) Phys. Med. Biol. 31, 1295–1297.
[47] Blackshaw, S. and Snyder, S.H. (1999) J. Neurosci. 19, 3681–3690.
[48] Palczewski, K. et al. (2000) Science 289, 739–745.